DOCUMENT RESUME

ED 155 759 CS 502 145

AUTHOR	Mead, Nancy A.
TITLE	Issues Related to Assessing Listening Ability.
PUB DATE	Mar 78
NOTE	18p.; Paper presented at the Annual Meeting of the American Educational Research Association (Toronto, Canada, March 1978)
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	*Audio Passive Laboratories; Communication Skills; Educational Assessment; Language Skills; *Listening; Listening Comprehension; *Listening Skills; *Listening Tests; Research Needs; *Research Projects
IDENTIFIERS	*National Assessment of Educational Progress; *Speech Communication Association

ABSTRACT

The National Assessment of Educational Progress (NAEP) and the Speech Communication Association (SCA) initiated a pilot study to test the feasibility of assessing speaking and listening skills. A pool of 56 items was developed and then field tested at four sites which represented a variety of national regions, of size and type of cities, and of racial and ethnic populations. There were significant differences between the responses of minority and nonminority students. A panel of speech communication experts hypothesized that minority students might have less specialized vocabulary knowledge, a lower tolerance for long materials perceived as boring, and less experience listening to the accents and speaking rates of white speakers. The results of the NAEP/SCA project suggest a clear need for further development and research in this area. In addition, three guidelines were developed for researchers in the area of listening ability: (1) focus on skills that are unique and central to listening, (2) use short, interesting listening stimuli, and (3) consider extraneous factors which might contribute to item bias; these suggestions are intended to aid researchers rather than act as a definitive guide. (JF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED155759

ISSUES RELATED TO ASSESSING LISTENING ABILITY

by:
Nancy A. Mead

National Assessment of Educational Progress
March 1978

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Nancy A. Mead

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Presented at the 1978 Annual Meeting of the American Educational Research Association (Toronto) as part of a symposium entitled:

SPEECH COMMUNICATION COMPETENCE: PERSPECTIVES AND ISSUES

Chairman: Robert J. Kibler
International Communication Association and
Florida State University

CS502 145

ISSUES RELATED TO ASSESSING LISTENING ABILITY[1]

Introduction

During the fifties and early sixties, there was a flurry of research and instructional development in the area of listening ability. Since that time there has been little activity. Devine's (1978) recent summary of listening research primarily restates the conclusions which were reached by earlier reviewers (Caffrey, 1955; Toussaint, 1960; Russell, 1964; Dixon, 1964; Ducker and Petrie, 1964; Devine, 1967; Keller, 1969; Barker, 1971; Weaver, 1972). However, reliance on established answers to questions about listening appears to be misguided.

The impetus for this paper on the issues related to assessing listening ability comes from rather disappointing results in a pilot effort to develop listening measures for the National Assessment of Educational Progress. Our complacency concerning the task did not prepare us for the problems we encountered. We felt that listening ability, unlike the other aspects of communication competence which we were trying to assess, was fairly well defined. Furthermore, we were encouraged by the fact that there already existed good models for assessing listening. However, the results of our pilot efforts proved to be less than satisfactory. In fact, they directed us to some major redevelopment and research in the area of listening. In this paper I will share the results of the pilot effort of assessing listening ability and indicate the implications of this experience for future measurement of listening ability.

---

[1] Paper presented at the American Educational Research Association Convention, Toronto, Canada, March, 1978.

Brief Review of Listening Research, Instruction and Assessment

As indicated in the introduction, the major contributions toward defining and studying listening ability were made during the fifties and early sixties. Most of this activity involved

1. defining listening ability, either as a unitary skill or series of subskills;
2. exploring the relationship between listening ability and other factors, primarily verbal ability, reading ability and motivation;
3. developing and evaluating listening instructional approaches; and
4. developing tests of listening ability.

The results of the research on listening have not produced a single, empirically based definition of listening ability (Devine, 1978). Instead, a series of descriptions have emerged, mostly developed by those involved in listening instruction and measurement. These definitions are primarily based on a logical analysis of the listening process. Many follow established descriptions of reading comprehension. However, they do not indicate a clear hierarchy or scope and sequence of skills.

The research related to the relationship of listening and other factors has substantiated a relatively high positive correlation between listening and general verbal ability. Crook (1957) found a correlation of .70 between listening comprehension and intelligence. Haberland (1959) found generally high positive correlations between various listening and verbal ability measures. Likewise, the relationship between listening and reading has been well established. Brown (1965) reported correlations ranging from .76 to .82 between listening and reading for fourth, fifth and sixth graders. Ducker (1965) reported an average correlation between listening and reading of .57. However, it is possible that the relationship between

listening and reading may be explained by the overlap of both these abilities with general verbal ability.

Empirical research also has generally supported the relationship between listening ability and motivation or interest. This may be explained by the fact that listening is related to the general process of learning and the relationship between motiviation and learning has been well established (Barker, 1972). However, several experiments have shown that the effects of interest did not greatly influence listening comprehension (Heath, 1952; Karraker, 1964). A listening test is a special case of motivation. Most students want to perform well on tests. Kelly (1967) argued that listening ability measured in a acknowledged test situation is different from listening measured under normal conditions.

In addition to research studies, numerous programs which teach listening skills have been developed, including several commercial packages (e.g., Dun Donnelley, 1973; Educational Development Laboratories, 1969; McGraw-Hill, 1969; Science Research Associates, n.d.). Reviews of many studies of listening instruction state that in many but not all cases these programs have been effective in improving listening ability (Devine, 1967, 1978; Weaver, 1972). However, inspection of the contents of these programs indicates that they differ greatly in terms of the skills that they cover. Furthermore, it appears that the effectiveness of these programs to some extent depends upon the match of the instructional objectives and the evaluation instruments.

The various efforts at research and instructional were complemented by the development of several listening ability tests. Two standardized listening tests, the Brown-Carlsen (Harcourt Brace Jovanovich, 1955) and the Sequential Test of Educational Progress (Educational Testing Service,

1957), were developed in the mid-fifties and have been widely used ever since. However, some evidence indicates that these tests of listening correlate as well with tests of verbal ability and reading ability as they do with one another (Kelly, 1965). Perhaps this result can be explained by the fact that these two tests do not cover the same set of listening subskills. Besides these two standardized listening tests, numerous others have been reported in thesis research. However, most of these measures have not been carefully tested for reliability or validity.

Most efforts related to listening have depended upon the generalization which surfaced from the corpus of research described above. It was based upon this evidence that we embarked on the development of items that measure listening ability for the National Assessment of Educational Progress.

NAEP/SCA Pilot Listening Assessment

In June of 1976 the National Assessment of Educational Progress (NAEP) and the Speech Communication Association (SCA) initiated a pilot study to test the feasibility of assessing speaking and listening skills (Mead, 1977a, 1977b). The products of this effort were intended for use in the National Assessment of Educational Progress, a national survey of student achievement with respect to important educational objectives, funded by the National Center for Education Statistics.

There are some important differences between National Assessment and standardized achievement testing programs. The items developed by NAEP measure specific objectives which are considered important by educators and content specialists. They do not constitute a test per se. The items are used to describe the accomplishments of nationally representative

groups of students. They are not used to differentiate levels of ability among individual students. Nevertheless, the task of developing listening assessment items was similar to standardized test development in that it involved defining the domain of listening ability and constructing items which measured that domain.

The domain description developed by the NAEP/SCA pilot project reflected a somewhat broader definition of listening than is typical. The major focus of this description was on the functions or purposes of communication. These were identified as the informing function and the controlling (or persuading) function. The functions were further differentiated by the context or setting of the listening task. These included formal and informal listening situations. Finally, the domain was defined in terms of specific listening skills or objectives. This included general listening comprehension objectives as well as specific listening analysis objectives.

Another way of describing the domain is by characterizing the listening stimuli and questions which were developed for the pilot project. Stimuli representing the informing function included an informative speech, a telephone call, a newscast, and a public service announcement. The stimuli representing the controlling function included a persuasive speech, a paid political announcement, a commercial, and a persuasive conversation between two friends. Some of the questions which followed the stimuli measured general listening comprehension, specifically simple recall, interpretation and application. Other questions measured specific analysis objectives. These items required identifying appropriate introductions and conclusions, organizational patterns, types of support material, types of persuasive appeals, fact-opinion distinctions, and uses of evidence.

A pool of fifty-six items was developed. The items were packaged into four test booklets, each representing approximately fifteen minutes of testing. The items were field tested in four sites which represented a variety of regions of the country, size and type of cities, and racial and ethnic populations. An average of 140 students responded to each set of items.

The items were analyzed using typical item analysis statistics. Item difficulty was indicated by the percent of students choosing each option. Item discrimination was indicated by the point biserial correlation between individuals choosing an option and their total test scores. In addition, the responses to each option were correlated with an external criterion which reflected the classification of the students as minority or nonminority. This added information allowed reviewers to identify items which received significantly different responses by minority and nonminority students.

It is important to emphasize the purpose of adding the external criterion which reflected the racial/ethnic background of students to the information base. The aim of this strategy was not to eliminate all items which differentiated between minority and nonminority students. It is possible that there are real differences between these two groups with respect to listening ability. The information was used to identify items which might be discriminating between minority and nonminority students for reasons other than listening skill. For example, an item might receive different responses because of the varying backgrounds, experiences, values or language styles of minority and nonminority students. We considered these factors to be extraneous to listening ability.

The results of field testing indicated no problems with respect to selecting items with appropriate difficulty level and discrimination power. Guidelines had been established for selecting items within the difficulty range of forty percent to eighty percent correct responses with an average of sixty percent (Stanley and Hopkins, 1972) and with a discrimination level of .30 (Harris, 1968). Because the purpose of National Assessment is not to build a test but to select items which measure specific objectives, these guidelines were merely suggestive and not crucial. Practically all of the items in the pool met the discrimination requirement and only about twenty percent of the items fell outside the proposed difficulty range.

The surprising result from tryouts was that the listening items, unlike the items for the other areas of communication competence which were field tested at the same time, showed a high number of significant point biserial correlations between responses of minority and nonminority students. Approximately one-half of the listening items demonstrated this characteristic. It must be emphasized that a significant correlation between the responses of minority and nonminority students (a relationship significantly different from zero) was not considered tantamount to item bias. There were a couple of reasons for reviewing the data cautiously. First, the tryout sites included two all minority schools. This made it possible that the distributions might include a concentration of minority students within a single option because of some unusual responses by the students in these schools. Secondly, a great number of correlations were reviewed, one for each foil of each item. Among these, there were bound to be some relationships due to chance (one out of twenty).

A significant correlation was considered a signal for further review. In some cases, the critique indicated possible sources of item bias, such as a typically white speech pattern in a listening stimuli which presented a persuasive conversation between two friends, and the item was dropped. In other cases, however, the review could not detect any problems and the item was retained. As indicated earlier, a significant correlation was not considered to be synonymous with item bias. However, the frequency of this characteristic and the marked difference between this set of items and the other sets of items (informing speaking, controlling speaking, ritualizing and sharing feelings) suggested a special problem.

A panel of speech communication experts reviewed the listening items and selected approximately one-half for use in the assessment. About one-third of the selected items reflected significantly different responses by minority and nonminority students. The consultants identified very few specific aspects of the listening items which they felt were indicative of item bias, such as the type of situation presented, the speech style used, or the values implied. However, they speculated a number of general characteristics of the items which might have tapped factors which were extraneous to measuring listening ability. These problems included:

1. the vocabulary level of the listening stimuli;
2. the length of the formal speeches;
3. the interest level of the listening stimuli;
4. the accent and rate of speech of the speakers on the stimulus tapes; and
5. the level of disruption in the classrooms.

They hypothesized that minority students might have less specialized vocabulary knowledge; a lower tolerance for long, boring materials; and less experience listening to the accents and rate of white speakers.

Furthermore, minority students might tend to be concentrated in schools where there were more disruptions in the classrooms and nearby environment.

An additional factor which might explain the results is varying levels of verbal ability of the minority and nonminority students in the tryout groups. If listening ability overlaps with verbal ability, as previous research indicates, it is possible that the results might be explained in terms of different levels of verbal ability. The field testing did not collect information about the verbal ability of the students. It is possible that the minority students selected for tryouts reflected an overall lower level of verbal ability than the nonminority students.

The outcome of the tryout phase of the pilot listening assessment was the identification of a problem, potential item bias, and no real data to substantiate or further elaborate the situation. A number of explanations of the results was proposed. However, these explanations were based on speculation and not on empirical evidence. The problem of minority bias had not been clearly articulated in past listening assessment efforts. The results of the NAEP/SCA pilot project suggested a clear need for further development and research.

Implications for Future Development of Listening Measures

The message of this paper is that it is not as easy to assess listening ability as perhaps we have been led to believe. The problem of differing responses between minority and nonminority students not only flagged a potential problem in item bias, but also reopened more general issues regarding listening ability. These include questions about the definition of the domain of listening ability and about its relationship

with other factors (especially with verbal ability and motivation). A successful measure of listening ability must deal with all of these issues.

Based on our experience in the NAEP/SCA pilot project, we identified several recommendations for listening assessment which address the questions listed above. These guidelines have been adopted for our own continuing development of listening assessment items and are also relevant for others interested in measuring this skill.

Recommendation 1: Focus on Skills Which Are Unique and Central to Listening

As indicated in the second section of this paper, there still is no common definition of listening ability. In selecting from alternatives, it seems appropriate to focus on skills which are unique and central to listening.

The skills which are unique to listening involve responding to oral language. Spoken language is different from written language in that it tends to be nonlinear, incomplete and redundant. It is ephemeral, it is accompanied by nonverbal communication, and it often takes place in an interactive situation. It therefore seems essential to utilize natural spoken language for listening stimuli. The reworking of reading tests into listening tests is inappropriate. It is less obvious how to deal with the nonverbal and interactive nature of oral communication in an assessment situation. Nonverbal signals tend to be subtle and individualistic and thus difficult to include in an assessment. Likewise, the give and take of normal speaking and listening are difficult to recreate in a test setting.

It is more difficult to identify the most central skills in the listening domain. In our present effort, we have identified five core

objectives. These reflect a compilation of the skills most often identified in instructional and assessment materials. They include the following:

1. be able to recall significant details;
2. be able to comprehend the main idea;
3. be able to draw inferences about the information (e.g., relationships, implications);
4. be able to make judgments concerning the speaker (e.g., intent, attitudes); and
5. be able to make judgments concerning the information (e.g., types of evidence, logic of arguments).

However, these objectives suffer from the same problems as earlier lists. They are based on logic, not empirical evidence.

A final concern in defining the domain of listening is the overlap between listening ability and verbal ability. Listening skill depends upon knowledge of vocabulary and the ability to manipulate verbal symbols. However, it is also clear that it is possible to function effectively in many listening situations with a fairly limited vocabulary and with basic cognitive skills. Kelly (1967, pp. 455-456) described this contradictory situation as follows:

> In testing situations some of the "best" listeners may be subjects with high mental ability who normally are relatively inattentive under non-test circumstances, and some of those who are "good listeners" under normal (non-test) conditions may do poorly in the test environment because they were handicapped by the inability to understand the difficult material frequently found in the tests of listening.

Since the aim of listening assessment is to focus on skills which are central to this domain, it seems inappropriate to use materials and items which tap into high levels of verbal ability merely to gain discrimination power in the measures. General verbal skills are less amenable to improvement through listening instruction and they are already the focus of other types of assessment measures.

In our current research we are trying to explore the overlap between listening and verbal ability. Listening stimuli have been classified according to their vocabulary level by using readability formulae. In addition, listening items are being field tested along with verbal ability measures. This way it will be possible to see which stimuli and which items are particularly related to verbal ability. These efforts will also help sort out possible explanations to differing responses among minority and nonminority students in the initial tryouts.

Recommendation 2: Use Short, Interesting Listening Stimuli

The purpose of the present assessment effort and most tests of listening to date is to measure listening ability under optimal conditions rather than typical listening behavior in actual situations. As indicated in section two, motivation plays a critical role in listening. In order to assess maximum ability, it therefore seems appropriate to make every effort to encourage students to try their hardest on the items. For many students the test situation itself is an adequate motivator. However, more and more students are reacting to the relevance of their school experiences, including testing.

One problem identified in the initial tryouts was that stimuli were quite long (each speech was six minutes) and that the contents were uninteresting to seventeen-year-olds. In the present development effort, every attempt is being made to use relatively short, interesting listening stimuli. Stimuli range from one-half minute to three minutes in length. Materials focus on topics which are generally popular among teenagers.

It is difficult to find listening materials which all students will find interesting. Some will enjoy sports, others will not. In some

testing situations, this problem has been countered by using uniformly boring materials. However, in the current situation it seems more important to spark interest, even if it means introducing topics which might not motivate all students to the same degree. In fact, by focusing on topics and listening situations which are common to most students (for example, school activities, friendships, television), it is possible to provide stimuli that are both interesting and universal.

Recommendation 3: Consider Extraneous Factors Which Might Contribute to Item Bias

The primary result of the tryouts for the NAEP/SCA pilot project was the identification of potential minority bias in items. This finding highlighted the need for special attention to this problem.

One technique for identifying item bias, which was used in the pilot project, was to review correlational data regarding the responses of minority and nonminority students. Although we have already indicated some problems in using this information, it appeared a useful tool for finding extraneous factors such as background, experience and values.

Another potential contributor to bias is the quality of the listening stimulus. It seems likely that students who are use to listening to nonstandard dialects or to languages other than English might be confused by the listening stimuli. One solution to this problem is asking the teacher to read the stimuli in the testing situation. This assumes that all students are used to listening to their own teacher and will not be confused by his or her dialect or other speech characteristics. However, the lack of regularity in this type of testing situation and the possibility of speech problems among some teachers (e.g., poor articulation) favor an alternative approach.

In our present research we are developing stimulus tapes which use network English as the mode for presenting listening material. This approach is based on the assumption that all students are used to watching and listening to television. This technique allows for a high degree of regularity in the testing process. Additional considerations must be given to the listening environment, assuring that the loudness and tone of the stimuli are adequate and distractions are minimized. These approaches whould minimize the extraneous factors which cause problems for minority students.

The recommendations discussed above are general and suggestive. They are presented primarily as an impetus for further development and research. The guidelines are meant to encourage those who are involved in listening research, instruction or testing to explore the area more definitively, rather than to rely on existing data. The recommendations emerge more from the subjective experience of the NAEP/SCA pilot listening assessment project than from concrete empirical findings. These guidelines must be subjected to careful study and research. This is the goal of current National Assessment activities which are continuing to explore the area of listening, and hopefully,our efforts will be amplified by others interested in listening.

REFERENCES

Barker, L. L. Listening Behavior. Englewood, New Jersey: Prentice-Hall, 1971.

Brown, C. T. "Three Studies of the Listening of Children." Speech Monographs, 32 (1965), 129-138.

Caffrey, J. "Auding." Review of Educational Research, (1955), 121-138.

Crook, F. E. "Interrelationship Among a Group of Language Arts Tests." Journal of Educational Research, 51 (1957), 305-311.

Devine, T. G. "Listening." Review of Educational Research, 37 (1967), 152-158.

Devine, T. G. "Listening: What Do We Know after Fifty Years of Research and Theorizing?" Journal of Reading (1978), 296-304.

Dixon, N. R. "Listening: Most Neglected of the Language Arts." Elementary English, 41 (1964), 285-288.

Ducker, S. "Listening and Reading." Elementary School Journal, 65 (1965), 321-329.

Ducker, S. and Petrie, C., Jr. "What We Know about Listening: Continuation of a Controversy." Journal of Communication, 14 (1964), 245-252.

Dun Donnelley. Alive and Listening. New York: Dun Donnelley, 1973.

Educational Development Laboratories. Listen and Think. Educational Development Laboratories, 1969.

Educational Testing Service. A Brief Cooperative Sequential Test of Educational Progress. Princeton, New Jersey: Educational Testing Service, 1957.

Haberland, J. A. "A Comparison of Listening Tests with Standardized Tests." Journal of Educational Research, 52 (1959), 299-302.

Harcourt Brace Jovanovich. Brown-Carlsen Listening Comprehension Test. New York: Harcourt Brace Jovanovich, 1955.

Harris, M. L. "Some Methodological Suggestions for Construction of an Objective Measuring Instrument." Madison, Wisconsin: Wisconsin Research and Development Center for Cognitive Learning, University of Wisconsin, 1968.

Heath, M. A. "A Study on Listening: The Relationship Between Interest, Educability, and Score on an Objective Examination Over the Factual Content of an Informative Speech." Speech Monographs, 19 (1952), 159-160.

Karraker, M. "An Evaluation of the Influence of Interest and 'Set' on Listening Effectiveness in the Basic Communication Class." Speech Monographs, 14 (1964), 245-246.

Keller, P. "Major Finding in Listening in the Past 10 Years." Journal of Communication, 10 (1969), 29-38.

Kelly, C. M. "An Investigation of the Construct Validity of Two Commercially Published Listening Tests." Speech Monographs, 32 (1965), 139-143.

Kelly, C. M. "Listening: Complex of Activities -- And a Unitary Skill?" Speech Monographs, 34 (1967), 455-465.

McGraw-Hill. Listening: A Programmed Approach. New York: McGraw-Hill, 1969.

Mead, N. "The Development of an Instrument for Assessing Functional Communication Competence of Seventeen Year Olds." Unpublished doctoral dissertation, University of Denver, Denver, Colorado, 1977 (a).

Mead, N. "Issues Related to National Assessment of Speaking and Listening Skills." Unpublished paper presented at the Speech Communication Association Annual Convention, Washington, D.C., 1977 (b).

Russell, D. H. "A Conspectus of Recent Research on Listening Ability." Elementary English, 41 (1964), 262-267.

Science Research Associates. Listening Language Laboratory. Chicago: Science Research Associates, n.d.

Stanley, J. C. and Hopkins, K. D. Educational and Psychological Measurement and Evaluation. Englewood Cliffs, New Jersey: Prentice-Hall, 1972.

Toussaint, I. "A Classified Summary of Listening -- 1950-1959." Journal of Communication, 10 (1960), 125-134.

Weaver, C. H. Human Listening. Indianapolis, Indiana: Bobbs-Merrill, 1972.